Full-model-free Adaptive Graph Deep Deterministic Policy Gradient Model for Multi-terminal Soft Open Point Voltage Control in Distribution Systems

Huayi Wu, Zhao Xu, Minghao Wang, and Youwei Jia

***Abstract*—High penetration of renewable energy sources (RESs) induces sharply-fluctuating feeder power, leading to voltage deviation in active distribution systems. To prevent voltage violations, multi-terminal soft open points (M-SOPs) have been integrated into the distribution systems to enhance voltage control flexibility. However, the M-SOP voltage control recalculated in real-time cannot adapt to the rapid fluctuations of photovoltaic (PV) power, fundamentally limiting the voltage controllability of M-SOPs. To address this issue, a full-model-free adaptive graph deep deterministic policy gradient (FAG-DDPG) model is proposed for M-SOP voltage control. Specifically, the attention-based adaptive graph convolutional network (AGCN) is leveraged to extract the complex correlation features of nodal information to improve the policy learning ability. Then, the AGCN-based surrogate model is trained to replace the power flow calculation to achieve model-free control. Furthermore, the deep deterministic policy gradient (DDPG) algorithm allows FAG-DDPG model to learn an optimal control strategy of M-SOP by continuous interactions with the AGCN-based surrogate model. Numerical tests have been performed on modified IEEE 33-node, 123-node, and a real 76-node distribution systems, which demonstrate the effectiveness and generalization ability of the proposed FAG-DDPG model.**

***Index Terms*—Soft open point, graph attention, graph convolutional network, reinforcement learning, deep deterministic policy gradient.**

I. Introduction

THE increasing deployment of renewable energy sources (RESs) has resulted in significant voltage violations in distribution systems, since the outputs of photovoltaics (PVs) and wind turbine (WT) generators fluctuate rapidly [1]. To meet the power quality requirement of the flexible operation and management in distribution systems, voltage regulation has been receiving increasing academic attentions recently.

Various var devices are deployed in distribution systems to mitigate the voltage deviations. Traditionally, the on-load tap changers (OLTCs) and capacitor banks (CBs) are employed to regulate the voltage in a slow-response manner, which can hardly cope with fast voltage variations caused by the rapid fluctuations of RESs [2]. With the development of advanced power electronic devices, the distribution system with multi-terminal soft open points (M-SOPs) can provide enhanced controllability to achieve better operation flexibility [3]. Specifically, the M-SOP bolsters power flow control by adeptly redirecting electricity between various feeders, thereby offering more accurate voltage control. With the ability to manage fluctuating power outputs, M-SOP can also facilitate the integration of RESs, particularly beneficial in improving the renewable energy penetration of distribution systems. Besides, M-SOPs introduce operation flexibility, seamlessly integrating with existing infrastructure to optimize the network performance. The M-SOP can achieve satisfactory voltage regulation via fast and accurate control of the active and reactive power flows [4]. Generally, the M-SOP has been considered as a promising method to address the voltage issues caused by the high penetration of intermittent RESs.

The uncertainty optimization programming methods are introduced to quantify the uncertainties of PV outputs. Typical methods include stochastic optimization [5] and robust optimization [6]. The stochastic optimization is reported in [7] to regulate nodal voltages against PV uncertainties, which describes the uncertainty relying on numerous scenarios generated by parametered probabilistic distribution function. However, the specific information of uncertainty is rarely tractable in practice. It also suffers from heavy computational bur-

Manuscript received: February 19, 2024; revised: March 26, 2024; accepted: May 8, 2024. Date of CrossCheck: May 8, 2024. Date of online publication: May 31, 2024.

This work was supported by the National Natural Science Foundation of China (No. 72331008), Guangdong Natural Science Foundation (No. 2023A1515010653), Environment and Conservation fund (No. ECF 49/2022), and PolyU research project 1-YXBL and CDAH.

H. Wu and Z. Xu (corresponding author) are with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China, and Z. Xu is also with Shenzhen Research Institute, The Hong Kong Polytechnic University, Shenzhen, China (e-mail: huayiwu@polyu.edu.hk; eezhaoxu@polyu.edu.hk).

M. Wang is with the Department of Electrical and Computer Engineering, University of Macau, Macau, China (e-mail: mhwang@um.edu.mo).

Y. Jia is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China (e-mail: jiayw@sustech.edu.cn).

DOI: 10.35833/MPCE.2024.000177

den. To address these issues, robust optimization is employed to reduce the reliance on accurate information of the probability of distributions [8]. The robust optimization achieves solution to the worst-case scenario for immunizing against all the variation scenarios generated by the predefined uncertainty set [9]. A two-time-scale robust optimization method is proposed in [10] for M-SOP scheduling to mitigate voltage imbalance. A robust optimization method involving second-order cone format of distribution system is introduced in [11] to achieve the optimal operation of soft open point (SOP). The distributionally robust chance-constrained method is introduced in [12] to optimize the SOP operation to mitigate the impact of the load altering attack. The model-based stochastic and robust optimizations assume that accurate parameters of distribution system are available, which can be rarely guaranteed in practice [13]. Besides, these methods cope with uncertainties of RESs and loads by searching a feasible solution with the predefined uncertain interval. However, RESs may fluctuate frequently caused by the quick cloud shadow transitions [14]. Therefore, the practice of re-computing the optimization for voltage control may not be adaptive to the rapidly changing outputs of PVs. A faster response voltage control strategy is required to immunize the violent volatility of PVs.

To address the issue of model parameter dependency, the data-driven methods are developed. A data-driven method establishing the sensibility related to voltage regulation solutions and objective is proposed in [15] for operation of SOP, which allows changing system operation states. However, the optimization calculation is too time-consuming in real-time implementation to adapt to the RES power fluctuations. The historical data of forecasting error are utilized in [16] to construct the histograms so as to estimate the true probability distribution functions of the RESs. The exact parameters of system are mandatory, which are rarely available in practice.

To alleviate the need for accurate and timely model parameters, model-free methods are proposed in recent years. Deep reinforcement learning (DRL) can be trained offline with iteration between the simulation model and environment, and is finally applied online to the real distribution system. Two-stage volt-var control combining the DRL and the physical model is proposed in [17]. The DRL is proposed in [18] to obtain the optimal electric vehicle (EV) charging strategy in the distribution system while satisfying all the physical constraints. The DRL-based adaptive voltage control model is proposed in [19] for M-SOP operation to mitigate voltage violations. These methods rely on the modelling of distribution system to calculate the reward during the training process and real-time application, which require accurate information of the power system. Besides, the neural network utilized by these methods fails to consider the distribution system structure, which cannot extract the nodal features efficiently. To bridge this gap, graph neural network (GNN) with powerful graphical learning representation ability is receiving trending attention. It is applied in other areas. The unrolled spatiotemporal neural network is proposed to achieve real-time state estimation, representing significant ability in grid representation [20]. Besides, to fully leverage graphical feature extraction merit, the graph convolutional network (GCN)-based DRL model is successfully applied in distribution system restoration [21] and dispatch [22]. However, to the best of the authors' knowledge, none of the existing studies on data-driven optimal operation of SOP consider the possible spatial associations of nodal features in power systems. Actually, RES uncertainties may occur in power flow congestion and voltage fluctuation that propagate to the entire system. Thus, the accurate perception of the spatial nodal data can lead to more effective SOP decision-making. This motivates the feature extraction of graphically correlated nodal data for more effective SOP solutions.

This paper proposes a full-model-free adaptive graph deep deterministic policy gradient (FAG-DDPG) model for M-SOP voltage control in distribution system with high penetration of RESs. Traditionally, the fully connected and convolutional neural networks (CNNs) employed in actor and critic functions of reinforcement learning models often overlook the graphical structure inherent in power systems, resulting in inadequate capture of the complex graph-based features emerging from power flow variability. Due to the strong capability of GCN in graph feature representation, the attention-based adaptive graph convolutional network (AGCN)-based surrogate model is proposed for graph feature representation in extracting the variations in power flow due to RES and M-SOP control, enabling improved policy learning and generalization capabilities for network reconfiguration. Besides, traditional methods rely on historical data and physical model-based calculations to obtain system states during the reinforcement learning process, which is computationally cumbersome. To address this issue, the AGCN-based surrogate model is learned to rapidly observe states based on measurements, facilitating the model-free voltage control. Furthermore, the proposed FAG-DDPG model iteratively learns policy through interactions with both the AGCN-based surrogate model and the environment. In this way, the continuous control variables of M-SOP can be achieved to mitigate voltage violations in real-time to deal with the rapid fluctuations of RESs. The contributions of this paper can be summarized as follows.

1) The propsoed FAG-DDPG model is proposed to train a model-free M-SOP control strategy, facilitating real-time and continuous decision-making based on the latest observations. In this way, voltage fluctuations due to rapid fluctuations of RESs can be efficiently mitigated.

2) Different from utilizing the topology-unaware neural network, the attention-based AGCN is proposed in the deep deterministic policy gradient (DDPG) algorithm to extract the graphical features representing the interrelations among nodes to achieve increased adaptability to new configurations.

3) Instead of utilizing the perfect physical model parameters, the AGCN-based surrogate model is coordinated with the DDPG to achieve a model-free M-SOP control to enhance the practicality and applicability of the model.

The remainder of this paper is organized as follows. The distribution system with M-SOP voltage control is proposed in Section II. The proposed FAG-DDPG model for M-SOP

voltage control is introduced in Section III. Numerical case studies are presented in Section IV. Finally, Section V concludes this paper.

II. Distribution System with M-SOP Voltage Control

A. Control Objective

Voltage control in distribution systems typically encompasses the deployment of OLTCs, CBs, static var compensators (SVCs), etc. OLTCs and CBs, characterized by their slower response dynamics, are generally scheduled on a day-ahead basis to accommodate their operation limitations [23]. The SVCs are popular with their fast and continuous response characteristics, but their deployment is often limited by high costs and their design is tailored on high-voltage power systems with significant reactive power needs. To circumvent these constraints, the M-SOP is implemented for intraday voltage regulation. This method facilitates the integration of day-ahead dispatch decisions for OLTCs and CBs, thereby enhancing the overall efficiency and responsiveness of distribution systems.

The control variables of M-SOP are active power and reactive power. By scheduling the active power transfer, the converters can alleviate the nearby power congestion on the power line. In this way, the local voltage profile near the converters can be improved. Thus, the total voltage deviation of each node is minimized and set as the control objective in the model as:

$$\min \sum_{i=1}^{N} |U_i - U_0| \tag{1}$$

where U_i and U_0 are the voltage magnitude at node i and the rated voltage magnitude, respectively; and N is the total number of nodes in the distribution system.

B. Operation Constraints of Distribution System

The DistFlow branch model [24] is introduced to model the distribution system operation. The constraints are formulated as:

$$P_i^{PV} + P_i^{SOP} - P_i^{L} = \sum_{k:j\to k} P_{jk} - \sum_{i:i\to j} (P_{ij} - r_{ij} l_{ij}) \quad \forall i \in N, ij \in E \tag{2}$$

$$Q_i^{PV} + Q_i^{SOP} - Q_i^{L} = \sum_{k:j\to k} Q_{jk} - \sum_{i:i\to j} (Q_{ij} - x_{ij} l_{ij}) \quad \forall i \in N, ij \in E \tag{3}$$

$$U_j^2 = U_i^2 - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2) I_{ij}^2 \quad \forall i \in N, ij \in E \tag{4}$$

$$(U_i^{\min})^2 \le U_i^2 \le (U_i^{\max})^2 \tag{5}$$

$$I_{ij}^2 \le (I_{ij}^{\max})^2 \tag{6}$$

where P_i^{PV} and P_i^{L} are the active power injected by PV outputs and load demands, respectively; P_{ij} and Q_{ij} are the active and reactive power from bus i to bus j, respectively; $l_{ij} = |I_{ij}|^2$; Q_i^{PV} and Q_i^{L} are the reactive power generated by PV outputs and load demands, respectively; P_i^{SOP} and Q_i^{SOP} are the active and reactive power injections from the i^{th} terminal of SOP to the connected points of the network, respectively; r_{ij} and x_{ij} are the resistance and reactance of line ij, respectively; E is the set of lines; I_{ij} is the current in line ij; $U_i^{\min}$ and $U_i^{\max}$ are the lower and upper bounds of voltage magnitudes at node i, respectively; and $I_{ij}^{\max}$ is the maximum limit of current at line ij.

C. Operation Constraints of M-SOP

M-SOP is generally deployed among feeders to alleviate power congestion. In comparison to the optimization decision based on conventional mechanical switch state [7], M-SOP controls the active and reactive power flows more accurately with lower cost due to the avoidance of frequent switch actions.

The control variables of M-SOP consist of active power and reactive power of multiple converters. The positive direction of M-SOP converters is to inject active power into the system. The operation constraints in the steady-state model of M-SOP can be formulated as [25]:

$$\sum_{i=1}^{N_s} (P_i^{SOP} + P_i^{SOP,L}) = 0 \tag{7}$$

$$P_i^{SOP,L} = A_i^{SOP} \sqrt{(P_i^{SOP})^2 + (Q_i^{SOP})^2} \tag{8}$$

$$\sqrt{(P_i^{SOP})^2 + (Q_i^{SOP})^2} \le S_i^{SOP} \tag{9}$$

where N_s is the number of SOP nodes; $P_i^{SOP,L}$ is the power loss of converter connected to node i; A_i^{SOP} is the power loss coefficient of converter connected to node i; and S_i^{SOP} is the apparent power capacity of converter connected to node i.

III. Proposed FAG-DDPG Model for M-SOP Voltage Control

The proposed FAG-DDPG model for M-SOP voltage control is formulated as a Markov decision process (MDP). Firstly, the attention-based AGCN is introduced, and then the DDPG is described, followed by its training process.

A. Graph Convolutional Operation

Conventionally, due to the strong fitting ability of the CNN and the fully connected network (FCN), they are introduced to the DRL model for voltage regulation in distribution system [26]. The distribution system can be abstracted as a graph, such that CNN and FCN may not adapt to data learning from a graph representation perspective. As a result, the approximating performance may be inefficiency. Besides, the GNN being applied to power flow solution illustrates its superiority in graphical data feature extraction to achieve power system representation [27]. Thus, GCN is chosen to model the FAG-DDPG functions.

The graphical structure of the distribution system can be represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2, \ldots, v_i, \ldots, v_N\}$ is the set of all N nodes; and $\mathcal{E}$ is the set of branches. The graph convolutional operation is defined as:

$$\boldsymbol{H} = \tilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W} \tag{10}$$

where $\boldsymbol{H}$ is the output; $\boldsymbol{X}$ is the input; $\boldsymbol{W}$ is the matrix of learning parameters; and $\tilde{\boldsymbol{A}}$ is the graph operator which is often defined by $\tilde{\boldsymbol{A}} = \boldsymbol{D}^{-\frac{1}{2}}\hat{\boldsymbol{A}}\boldsymbol{D}^{-\frac{1}{2}}$, $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}_N$, $\boldsymbol{A}$ is the adjacent matrix, $\boldsymbol{D}$ is the degree matrix with its elements defined as $D_{ii} = \sum_j \hat{A}_{ij}$, and $\boldsymbol{I}_N$ is the identity matrix.

B. Operation of AGCN-based Surrogate Model

Generally, the adjacency matrix corresponding to the structure of the distribution network is introduced to design the graph operator $\tilde{\boldsymbol{A}}$ as aforementioned. Such graph operator with fixed values as its elements cannot extract the implicit dynamic and violently uncertain features hidden in the input data, and thus leading to unsatisfying strategy learning performance. To address this issue, the attention mechanism that can achieve dynamic graphical data feature extraction motivates the proposal of adaptive graph operator. Specifically, the incorporation of an attention mechanism within the model facilitates a focused analysis of pivotal nodes and edges in the system through the automated weighting of node interconnections. This method significantly augments the capability to discern and comprehend essential topological associations, which are vital for efficacious voltage regulation. Consequently, this enhancement empowers the model to adapt more proficiently to alterations in the network configuration. The attention mechanism can be described as a mapping process, where the embedding is transformed into query $\boldsymbol{Q}$, key $\boldsymbol{K}$, value $\boldsymbol{V}$ pairs, resulting in an output, which are formulated as:

$$\begin{cases} \boldsymbol{Q}=\boldsymbol{W}^{Q}\boldsymbol{X} \\ \boldsymbol{K}=\boldsymbol{W}^{K}\boldsymbol{X} \\ \boldsymbol{V}=\boldsymbol{W}^{V}\boldsymbol{X} \end{cases} \tag{11}$$

where $\boldsymbol{W}^{Q}$, $\boldsymbol{W}^{K}$, and $\boldsymbol{W}^{V}$ are trainable weights for query $\boldsymbol{Q}$, key $\boldsymbol{K}$, and value $\boldsymbol{V}$, respectively.

The softmax of the inner product $\boldsymbol{Q}$ and $\boldsymbol{K}$ represents the dependence weights among different nodal information. It serves as the score to compute the weighted sum of values by performing a linear transformer.

$$\boldsymbol{A}_{atten}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{12}$$

where d_k is the dimension of the matrix $\boldsymbol{K}$.

To extract the observation information from different representation subspaces, multiple-head operation is introduced:

$$head_i=\boldsymbol{A}_{atten}(\boldsymbol{W}_i^{Q}\boldsymbol{X},\boldsymbol{W}_i^{K}\boldsymbol{X},\boldsymbol{W}_i^{V}\boldsymbol{X}) \tag{13}$$

$$\boldsymbol{A}_{adp}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=Concat(head_1,head_2,\ldots,head_i)\boldsymbol{W}_{con} \tag{14}$$

where $\boldsymbol{W}_i^{Q}$, $\boldsymbol{W}_i^{K}$, and $\boldsymbol{W}_i^{V}$ are the projection parameters; *Concat* is the concatenation operation; $\boldsymbol{W}_{con}$ is the trainable parameter; and $\boldsymbol{A}_{adp}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})$ is the output of concatenation operation.

C. Formulation of MDP

The decision-making process for M-SOP control can be formulated as an MDP with finite time steps. The MDP is often described by tuples $\{\mathcal{S},\mathcal{A},\mathcal{T},\mathcal{R}\}$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{T}$ is the transition space, and $\mathcal{R}$ is the reward space. The action space $\mathcal{A}$ consists of the active and reactive power outputs from converters, i.e., $a_{i,t}=\{P_{1,t}^{SOP},\ldots,P_{N_s-1,t}^{SOP},\ Q_{1,t}^{SOP},\ldots,Q_{N_s,t}^{SOP}\},\forall i\in N_s,t\in T$, which are generated by the outputs of actor network. The reactive and active power outputs are limited by their minimum and maximum bounds, i.e., $Q_{i,\min}^{SOP}\le Q_i^{SOP}\le Q_{i,\max}^{SOP}$, $P_{i,\min}^{SOP}\le P_i^{SOP}\le P_{i,\max}^{SOP}$, to ensure that constraint (9) is satisfied, where $P_{i,\max}^{SOP}$ and $Q_{i,\max}^{SOP}$ are the maximum active power and reactive power, respectively; and $P_{i,\min}^{SOP}$ and $Q_{i,\min}^{SOP}$ are the minimum active power and reactive power, respectively. Since the SOP loss can be simplified to a suitable value [28], the loss value at the i^{th} terminal is calculated as $P_i^{SOP,L}=A_i^{SOP}S_{i,fix}^{SOP}$, $S_{i,fix}^{SOP}=\sqrt{(P_{i,\max}^{SOP})^2+(Q_{i,\max}^{SOP})^2}$, where $S_{i,fix}^{SOP}$ is the power capacity of converter connected to node i. Then, the active power output of the last terminal can be determined by $P_{N_s}^{SOP}=-\sum_{i=1}^{N_s-1}P_i^{SOP}-\sum_{i=1}^{N_s}P_i^{SOP,L}$. The state space $\mathcal{S}$ includes the PV power injections, load demands, and voltage profiles, which is defined as: $s_{i,t}=\{P_{i,t}^{L}-P_{i,t}^{PV},Q_{i,t}^{L}-Q_{i,t}^{PV}\},\forall i\in N,t\in T$, where $s_{i,t}$ is the state at node i and time t.

r_t is the reward function at time t, which is defined as:

$$\begin{cases} r_t=-\sum_{i=1}^{N}\left|\dfrac{U_i-U_0}{U_0}\right|+\eta\psi \\ \psi=|\max\{0,U_i-U_{\max}\}|+|\min\{0,U_i-U_{\min}\}| \end{cases} \tag{15}$$

where ψ is the penalty item representing the extent of voltage limit violation; η is the penalty coefficient; U_0 is the nominal voltage value usually set to be 1; and $U_{\max}$ and $U_{\min}$ are the maximum and minimum voltage limits, respectively.

The proposed FAG-DDPG model for solving the MDP consists of two main functions, i.e., the action-value function and the policy function. Figure 1 shows the iteration process of the proposed FAG-DDPG model.

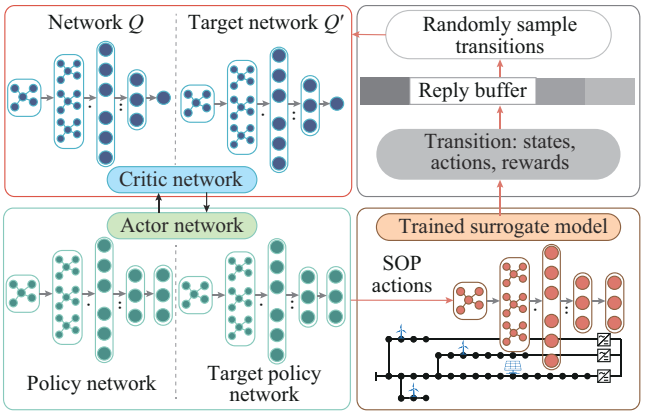

Fig. 1. Iteration process of proposed FAG-DDPG model.

D. Solution Method of Proposed FAG-DDPG Model

1) Policy Network

The actor network u is the policy function that maps states s_t to actions a_t so that the action for control strategy can be formulated as:

$$a_t=u(s_t|\theta^{u})+\mathcal{N}(0,\sigma_t) \tag{16}$$

where θ^{u} is the parameter of the actor network; $\mathcal{N}$ is the noise following normal distribution added in the actions generated by the actor network to conduct action exploration function; and σ_t is the standard deviation. This stochastic perturbation is critical for ensuring sufficient exploration of the action space during the learning process, thereby preventing the model from prematurely converging to suboptimal policies and enabling it to discover more effective strategies

over time. Since the control model is designed to mitigate the voltage violation in practice, the action-value function is constructed to minimize the control objective Q as:

$$Q(s_t, a_t) = \sum_{t=1}^{T} [\gamma_t r_t | s_0, a_0] \tag{17}$$

where γ_t is the discounted factor at time t. $T = 1440$ is the total time horizon for each episode. To enhance the stability, at each process of action-environment iteration, a transition is generated and stored in the memory replay buffer $\mathcal{D}$, which is then sampled in batches to train the actor and critic networks. To maximize the expectation value $E_{\mathcal{D}}$ of the action-value function Q, the loss function J is formulated as:

$$J = \max_{\theta^u} E_{\mathcal{D}}[Q(s_t, a_t)] \tag{18}$$

This loss function is optimized by the gradient ascent method to update the parameters θ^u of the actor network. The value of J is maximized by giving more fitting actions a_t generated by the actor network. Thus, according to the chain rule, the gradient of J with respect to θ^u can be calculated by multiplying the gradient of action value Q with respect to the actions and the action a_t with respect to parameters θ^u in the actor network.

$$\nabla_{\theta^u} J = \frac{1}{m} \sum_{s_t \in \mathcal{D}} \nabla_{\theta^u} a_t(s_t|\theta^u) \nabla_{a_t} Q(s_t, a_t|\theta^Q) \tag{19}$$

where m is the mini-batch sampled in the memory replay buffer $\mathcal{D}$; and θ^Q is the parameter of the critic network.

2) Value Network

The critic network is the action-value function to calculate $Q(s_t, a_t)$ that maps (s_t, a_t) to a scalar, which is utilized to evaluate the voltage control effect by implementing the M-SOP control strategy a_t under the environment s_t. The critic network parameters are trained by the defined loss function L as:

$$L(\theta^Q) = E_Q[(Q(s_t, a_t|\theta^Q) - y_t)^2] \tag{20}$$

$$y_t = r(s_t, a_t) + \gamma_t Q(s_{t+1}, u(s_{t+1})|\theta^Q) \tag{21}$$

The critic network $Q(s_t, a_t|\theta^Q)$ is trained to get close to the label y_t.

3) Target Network

To improve the stability of learning process, the target actor network u' and target critic network Q' are introduced. They are the copies of the original actor and critic networks by storing the earlier trained parameters. In this way, the label y_t for action-value function can be reformulated by the target networks as:

$$y_t = r(s_t, a_t) + \gamma Q'(s_{t+1}, u'(s_{t+1})|\theta^{Q'}) \tag{22}$$

The parameters of target networks $\theta^{u'}$, $\theta^{Q'}$ are updated by slowly tracking the online learned network with weighted coefficient τ.

$$\begin{cases} \theta^{u'} \leftarrow \tau\theta^u + (1-\tau)\theta^{u'} \\ \theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \end{cases} \tag{23}$$

E. AGCN

The state space data are composed of the inputs of networks in the proposed FAG-DDPG model, where the PV power represents complex spatial correlations. Therefore, the AGCN is introduced to construct the policy function and value function networks. In this way, the correlation features among PVs can be extracted to facilitate the learning performance. Note that the AGCN and FCN are employed to structure these networks to learn the policy and Q value. Specifically, the relationship details between the inputs and outputs of these networks are introduced as follows.

The inputs of the policy network are the states. The outputs are the SOP dispatch decisions. The AGCN and FCN are employed to approximate the policy function $a_t = u(s_t)$. The AGCN layers are utilized to embed the correlation features into the latent features that are mapped into the action space by the following FCN layers.

$$\begin{cases} a_t = tanh(\boldsymbol{H}_{out}^u) \\ \boldsymbol{H}_l^{u,fcn} = ReLU(\boldsymbol{H}_{l-1}^{u,fcn} \boldsymbol{W}_l^{u,fcn} + \boldsymbol{b}_l^{u,fcn}) & l \in \Omega^{Lfcn} \\ \boldsymbol{H}_l^{u,gcn} = \sigma(\boldsymbol{A}_{adp} \boldsymbol{H}_{l-1}^{u,gcn} \boldsymbol{W}_l^{u,gcn} + \boldsymbol{b}_l^{u,gcn}) & l \in \Omega^{Lgcn} \\ \boldsymbol{H}_0^u = \{s_t\} \end{cases} \tag{24}$$

where $\boldsymbol{H}_0^u$, $\boldsymbol{H}_l^{u,gcn}$, $\boldsymbol{b}_l^{u,fcn}$, and $\boldsymbol{H}_{out}^u$ are the outputs of the input layer, the AGCN layers, the FCN layers, and the output layer in the actor network, respectively; $\boldsymbol{W}_l^{u,gcn}$ and $\boldsymbol{W}_l^{u,fcn}$ are the parameters of the AGCN layers and the FCN layers, respectively; $\boldsymbol{b}_l^{u,gcn}$ and $\boldsymbol{b}_l^{u,fcn}$ are the corresponding bias parameters; $\sigma(x) = 1/(1+e^{-x})$ is the sigmoid activation function; $ReLU(\cdot)$ is the rectified linear unit function; $tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$ is the activation function, which limits the output layer values to a specified interval $[-1, 1]$; and Ω^{Lfcn} and Ω^{Lgcn} are the sets of FCN and AGCN layers, *Lfcn* and *Lgcn* denote the numbers of FCN and AGCN layers, respectively. The trainable parameters $\{\boldsymbol{W}_l^{u,gcn}, \boldsymbol{W}_l^{u,fcn}, \boldsymbol{b}_l^{u,gcn}, \boldsymbol{b}_l^{u,fcn}\}$ are grouped into θ^u. In this way, the action outputs can be ensured to satisfy the specified constraints without exceeding them. Besides, the Q network has the same structure of the policy network except for the output size of 1.

F. AGCN-based Surrogate Model

The low-voltage distribution system parameters are difficult to obtain in practice due to the high complexity to estimate them. This problem spurs the motivation to build up an AGCN-based surrogate model $f^s(\cdot)$ of the mapping relationship between the power injections and the voltage magnitudes. In this way, the proposed AGCN-based surrogate model serves as a computationally streamlined substitute for conventional power flow simulations, enabling swift reward acquisition throughout the training phase. Besides, this AGCN-based surrogate model obviates the need for precise network parameters, thereby facilitating a model-free method. Since the GNN model applied in power flow shows significant advantage on power system representation, the operations of AGCN and FCN are leveraged to learn the surrogate mode. By introducing the adaptive graph operator, it can extract the high-dimensional features from the training data. The inputs of the AGCN-based surrogate model consist of net active power injections $P_i = P_i^{PV} - P_i^L$ and net reactive power injections $Q_i = Q_i^{PV} - Q_i^L$ at each node. The outputs are the voltage magnitudes U_i at each node. The formulations of this model are expressed as:

$$\begin{cases} f^s = ReLU(\boldsymbol{H}_{out}^s) \\ \boldsymbol{H}_l^{s,fcn} = ReLU(\boldsymbol{H}_{l-1}^{s,fcn}\boldsymbol{W}_l^{s,fcn} + \boldsymbol{b}_l^{s,fcn}) & l \in \Omega^{Lfcn} \\ \boldsymbol{H}_l^{s,gcn} = \sigma(\boldsymbol{A}_{adp}\boldsymbol{H}_{l-1}^{s,gcn}\boldsymbol{W}_l^{s,gcn} + \boldsymbol{b}_l^{s,gcn}) & l \in \Omega^{Lgcn} \\ \boldsymbol{H}_0^s = [P_i, Q_i] \end{cases} \tag{25}$$

where $\boldsymbol{H}_0^s$, $\boldsymbol{H}_l^{s,gcn}$, $\boldsymbol{H}_l^{s,fcn}$, and $\boldsymbol{H}_{out}^s$ are the outputs of the input layer, the GCN layers, the FCN layers, and the output layer in the AGCN-based surrogate model, respectively; $\boldsymbol{W}_l^{s,gcn}$ and $\boldsymbol{W}_l^{s,fcn}$ are the parameters of the GCN layers and the FCN layers, respectively; and $\boldsymbol{b}_l^{s,gcn}$ and $\boldsymbol{b}_l^{s,fcn}$ are the corresponding bias parameters. $\theta^s = \{\boldsymbol{W}_l^{s,gcn}, \boldsymbol{W}_l^{s,fcn}, \boldsymbol{b}_l^{s,gcn}, \boldsymbol{b}_l^{s,fcn}\}$ is the trainable parameter of AGCN-based surrogate model.

G. Training Procedure

The training procedure of the proposed model composes of two stages. The first stage is to train the AGCN-based surrogate model. The loss function of the surrogate model (L1-loss) is defined as:

$$Loss(\theta^s) = \frac{1}{B_s}\sum_{b_s=1}^{B_s}|U_{b_s} - f_{b_s}^s| \tag{26}$$

where B_s is the batch size; U_{b_s} is the ground truth voltage magnitude; and $f_{b_s}^s$ is the output of surrogate model. The second stage is the training of proposed FAG-DDPG model, which is given in Algorithm 1.

Algorithm 1: training of proposed FAG-DDPG model

Randomly initialize the parameters of the critic network $Q(s,a|\theta^Q)$ and actor network $u(s|\theta^u)$. Initialize target networks Q' and u' with parameters $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{u'} \leftarrow \theta^u$. Initialize the memory replay buffer $\mathcal{D}$
For *episode* = 1 to K **do**
  Initialize a random process for action exploration
  Receive initial observation state s_1
  For $t = 1$ to T **do**
    Chose control actions $a_t = u(s_t|\theta^u) + \mathcal{N}(0, \sigma_t)$
    Apply action a_t, observe reward r_t, and state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$
    Sample a batch for m transitions from $\mathcal{D}$
    Update critic based on (20) and (21)
    Update actor policy based on (19)
    Update target networks based on (23)
  End for
End for

IV. Case Study

The schematic of the modified IEEE 33-node distribution system [29] with a three-terminal SOP is shown in Fig. 2. Four PVs are integrated into the system at nodes 12, 20, 24, and 28. The active power output of each PV is set to be 0.6 MW. The upper and lower bounds of the voltage are set to be 1.05 and 0.95 p. u., respectively. The proposed FAG-DDPG model consists of an AGCN-based surrogate model and a control model. The AGCN-based surrogate model learns the mapping from the power injections (input) to the nodal voltages (output) in a supervised manner. To achieve this goal, the dataset is generated by the following rules. The PV outputs are generated following the normal distribution with ±10% of the expected values as the standard deviation. The load demands are generated following the uniform distribution with interval [0.8, 1.2]. The curves of WT outputs, PV outputs, and load demand are shown in Fig. 3. Then, the AC power flow are calculated under these scenarios via MATPOWER program. After these processes, 50000 samples are generated, which are randomly divided into two parts, i. e., 90% are training data and 10% are testing data. The parameters of AGCN-based surrogate model are listed in Table I.

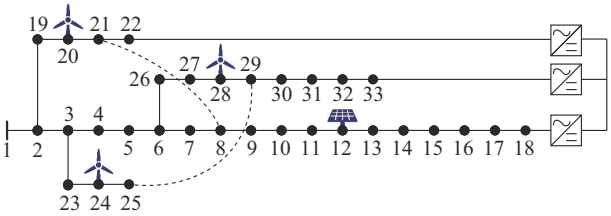

Fig. 2. Schematic of modified IEEE 33-node distribution system.

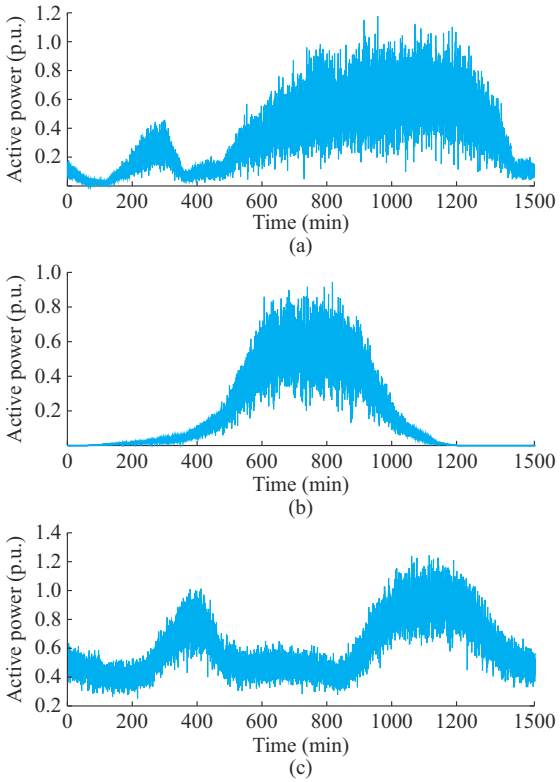

Fig. 3. Curves of WT outputs, PV outputs, and load demand. (a) WT outputs. (b) PV outputs. (c) Load demand.

TABLE I
Parameters of AGCN-based Surrogate Model

Parameter	Value
Dimension size of matrix $\boldsymbol{K}$	10
Neuron numbers of hidden layers	800/400
Learning rate	0.001
Batch size	1000
The maximum training epochs	10000

For the reinforcement learning based control model, the training data consist of 60 days' and 1 day' data, respective-

ly, with 1440 min in each day. The main architecture of actor and critic networks are set the same. The parameter settings of the adaptive graph reinforcement learning are listed in Table II. These parameters are set according to the optimization results by random searching selection. The experimental simulation is conducted by Python/Gurobi on a workstation with NVIDIA GeForce GTX 3090 GPU with 24 GB RAM.

TABLE II
PARAMETER SETTINGS OF ADAPTIVE GRAPH REINFORCEMENT LEARNING

Parameter	Value	Parameter	Value
Dimension size of matrix $\boldsymbol{K}$	10	Batch size	240
Neuron numbers of hidden layers	400/200	Step size of each episode	1440
Learning rate	0.001	Episode	4000
Memory replay buffer size	10000	Discount factor	0.9

A. Performance of AGCN-based Surrogate Model

The AGCN-based surrogate model is trained for 10000 epochs. The L1-loss proposed in (26) is to guide the training process. The loss values for each epoch during the training process is shown in Fig. 4(a). The logarithmic scale is utilized for better visualization. It is obvious that the loss gradually decreases with the increase of the epochs. This demonstrates that the proposed AGCN-based surrogate model can extract the dynamic graphical features in the measurements and learn the mapping from the power injections to the nodal voltages.

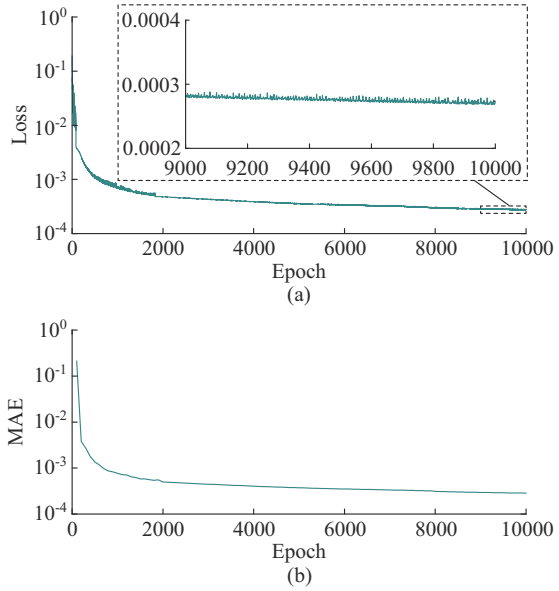

Fig. 4. Curves of loss and mean absolute error (MAE) during training process. (a) Loss of AGCN-based surrogate model. (b) MAE of testing data.

Furthermore, to verify the performance of the AGCN-based surrogate model on testing dataset, the MAE values of the testing data are depicted in Fig. 4(b). Notably, $MAE = \frac{1}{MN}\sum_{m=1}^{M}\sum_{i=1}^{N}|U_{m,i}-\hat{U}_{m,i}|$ is utilized as the evaluation index, where M is the total number of samples in the testing dataset; and $U_{m,i}$ and $\hat{U}_{m,i}$ are the actual and predicted voltage magnitudes, respectively. MAE can reflect the absolute distance between the predicted values and the ground truth values. The AGCN-based surrogate model is evaluated on testing data every 100 epochs. The MAE is 0.22 p.u. at the beginning, which is relatively high. With the increase of epochs, the MAE drops below 0.0003, demonstrating that the predicted voltages are every close to the true values. Overall, this curve has the consistent trend with the loss, indicating the effectiveness of the AGCN-based surrogate model.

Besides, to intuitively represent the performance of the AGCN-based surrogate model, the maximum and minimum voltage magnitude profiles calculated by the power flow and the AGCN-based surrogate model are depicted in Fig. 5. It shows that the voltage profile curves obtained by the power flow and the AGCN-based surrogate model are consistent with each other, which indicates that the proposed AGCN-based surrogate model can deliver high accuracy in voltage calculation. Therefore, the AGCN-based surrogate model can be leveraged to replace the power flow calculation to accelerate the training process of the FAG-DDPG model.

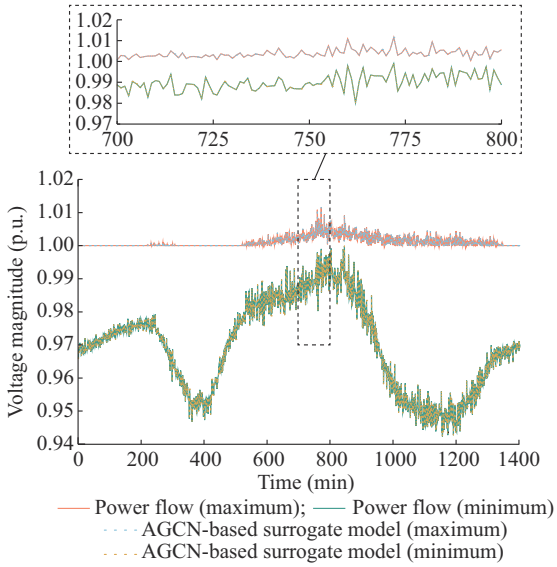

Fig. 5. The maximum and minimum voltage profiles calculated by power flow and AGCN-based surrogate model.

B. Performance of Proposed FAG-DDPG Model

1) Comparison with Traditional Optimization Method

To investigate the effectiveness of the proposed FAG-DRL, the second-order cone programing (SOCP) is conducted to optimize the SOP dispatch decision for every minute in a day. The objective function of the SOCP is set according to the reward. The absolute format for the objective function is transferred into linear format by linearization technique, i.e., $|X-Y|$ can be transferred to $A=|X-Y|$, $A \geq X-Y$, and $A \geq Y-X$. Thus, the SOCP is set as the baseline method. Three case settings are set as follows. Note that all the cases are based on the testing data.

1) Case I: there is no M-SOP voltage control, where the original voltage magnitude profiles are obtained.

2) Case II: the SOCP optimization is employed to achieve the optimal voltage control of M-SOP.

3) Case III: the proposed FAG-DDPG model is leveraged to obtain control action decisions of M-SOP.

The maximum and minimum voltage magnitudes in differ-

ent cases are depicted in Fig. 6. The specific voltage magnitudes of node 16 during the day in different cases are shown in Fig. 7. The voltage profiles of all nodes during the day in different cases are shown in Fig. 8. In case I, without any control action, the minimum voltages in some minutes are below the lower bound limit during the heavy load period. On the contrary, with the reactive and active power dispatch rescheduled by M-SOP employed in the feeder terminals, most of the minimum voltage magnitudes during the day are above 0.98 p.u. and below 1.02 p.u.. In Fig. 6, from 1100 to 1200 min, in comparison to the proposed FAG-DDPG model, the minimum voltage obtained by SOCP optimization methods fluctuates more violently. Figure 8 shows that all voltage profiles are improved in cases II and III. This indicates that the SOP can alleviate the voltage deviation via rescheduling the line power flow. Besides, in comparison with optimization method in case II, the maximum and minimum voltage profiles in case III obtained by proposed FAG-DDPG model have near optimal performance. Therefore, the proposed FAG-DDPG model can provide effective dispatch decisions to M-SOP to avoid voltage violations.

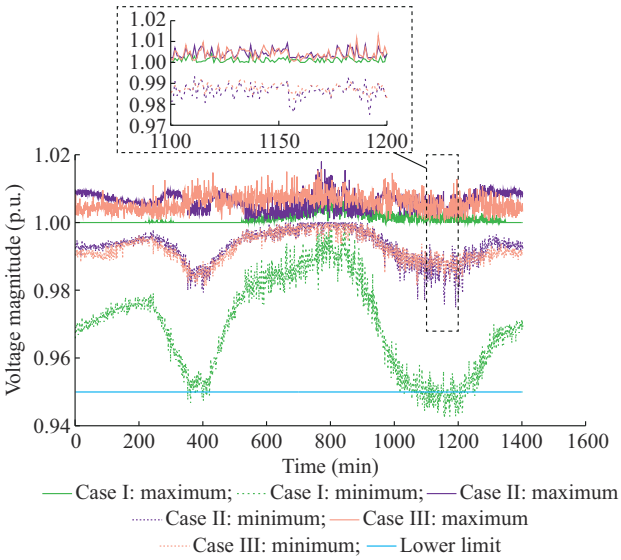

Fig. 6. The maximum and minimum voltage magnitudes in different cases.

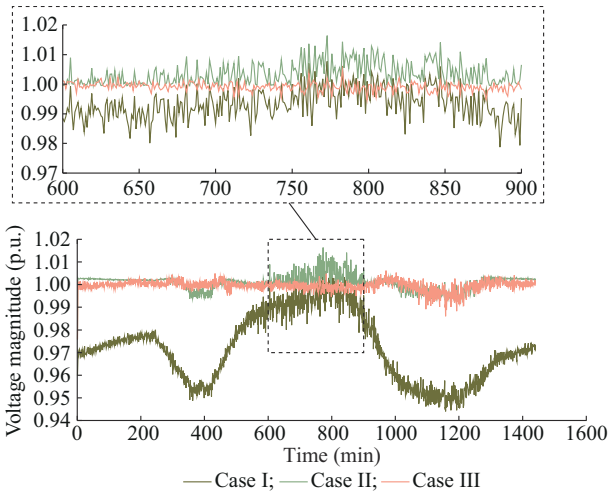

Fig. 7. Specific voltage profiles of node 16 during day in different cases.

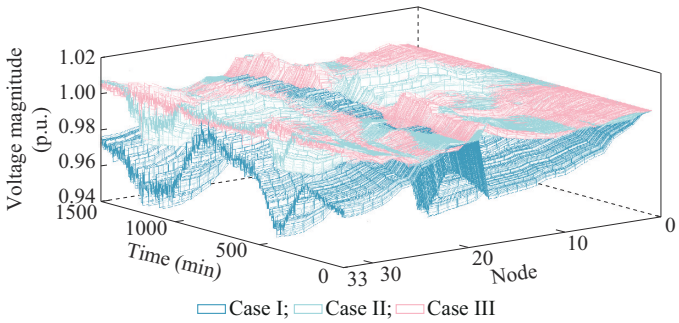

Fig. 8. Voltage profiles of all nodes during day in different cases.

2) Comparison with Other Deep Learning Based Models

To investigate the performance of the proposed FAG-DDPG model, the FCN, GCN, and CNN are utilized to replace the proposed model, which are defined as FCN-DDPG, GCN-DDPG, and CNN-DDPG, respectively. Then, they are employed as the baseline models. Note that the actor networks and critic networks are with the same deep learning structure. The trained AGCN-based surrogate model is shared for all agents. The FCN-DDPG model has three fully connected layers with 256, 128, and 64 neurons in each layer. The GCN-DDPG model includes a graph convolutional layer and three fully connected layers. The corresponding graph operator is the adjacency matrix calculated based on the distribution network topology. The two-dimensional (2D) CNN consists of three convolutional layers and three fully connected layers, where the channels are 16, 32, and 8, and the kernel sizes are 5.

The reward convergence curves during the training process for different deep learning based models are represented in Fig. 9. It is obvious that the proposed FAG-DDPG model achieves better reward than others. Besides, the convergence speed of the proposed FAG-DDPG model is faster than others.

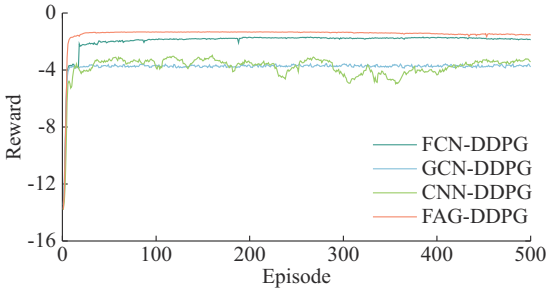

Fig. 9. Reward convergence curves during training process for different deep learning based models.

The efficacy of the proposed FAG-DDPG model is attributed to its adaptive graph operation, which proficiently captures latent graphical structure features in power distribution environments using its graph representation capabilities. Consequently, this enables the proposed FAG-DDPG model to efficiently learn voltage regulation strategies throughout the training process.

The maximum and minimum voltage magnitudes among all nodes during the day in every minute for different deep

learning based models are depicted in Fig. 10. It shows that most of the maximum and minimum voltages are below 1.04 p. u. and above 0.96 p. u.. However, compared with FCN-DDPG, GCN-DDPG, and CNN-DDPG, the maximum and minimum voltages obtained by the proposed FAG-DDPG model are within the interval between 1.02 p. u. and 0.98 p. u.. The fluctuations of RESs in distribution systems create a challenging learning environment due to significant fluctuations in nodal power injections. Traditional models such as FCN-DDPG, GCN-DDPG, and CNN-DDPG, as evidenced by lower rewards and higher voltage deviations, struggle in this complex setting. This is attributed to the lack of a graphical learning structure in FCN and CNN, which leads to an oversimplified treatment of nodal features without considering their interrelationships. Besides, while GCN can accommodate graphical data structures, its fixed graph operator is not well-suited for environments with frequent changes. In contrast, the higher rewards and more stable voltage profiles achieved by the proposed FAG-DDPG model indicate its superior capability in voltage regulation policy. This effectiveness stems from the attention mechanism of the AGCN-based surrogate model, which adeptly extracts and represents complex graphical features, thereby enhancing the mapping from environmental variables, characterized as power injections, to actionable decisions. Therefore, the proposed FAG-DDPG model can achieve graphical feature representation of the distribution system to ensure the effective voltage regulation policy against rapid fluctuations of RESs.

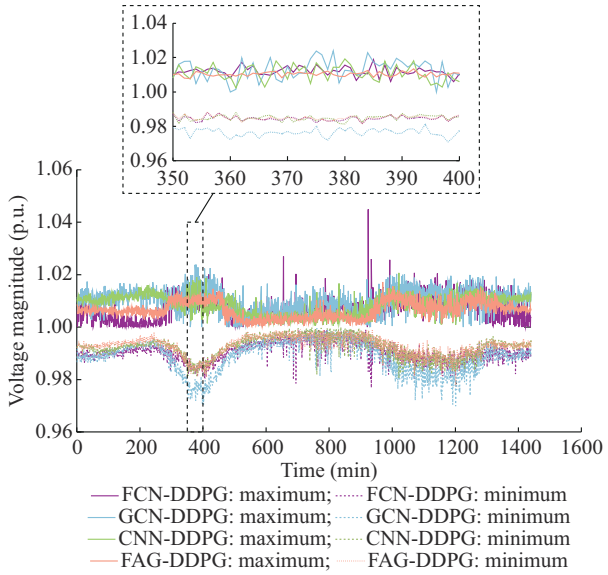

Fig. 10. The maximum and minimum voltage magnitudes for different deep learning based models.

The voltage profiles of node 16 for different deep learning based models are shown in Fig. 11. The voltage profiles of all nodes for different deep learning based models are characterized as mesh picture in Fig. 12. It can be observed that the voltage profile obtained by the proposed FAG-DDPG is closer to the reference voltage value with more moderate deviation. Besides, Table III summarizes the voltage deviations for different deep learning based models, which indicates that the proposed FAG-DDPG model can achieve the smallest mean voltage deviation. These results demonstrate that the proposed FAG-DDPG model can effectively learn the voltage control against fluctuations of RESs.

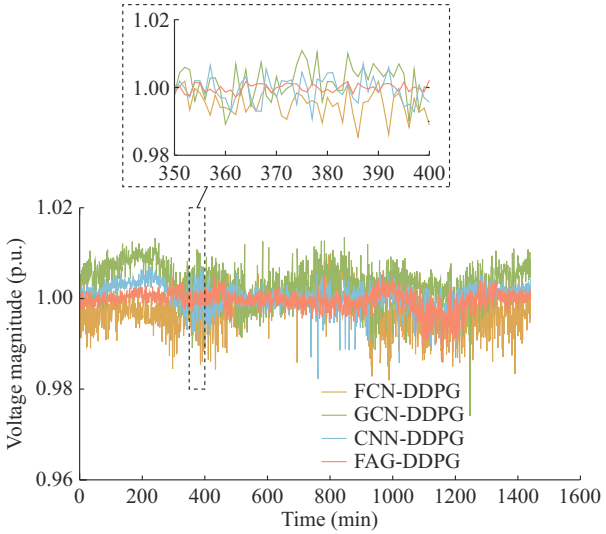

Fig. 11. Voltage profiles of node 16 for different deep learning based models.

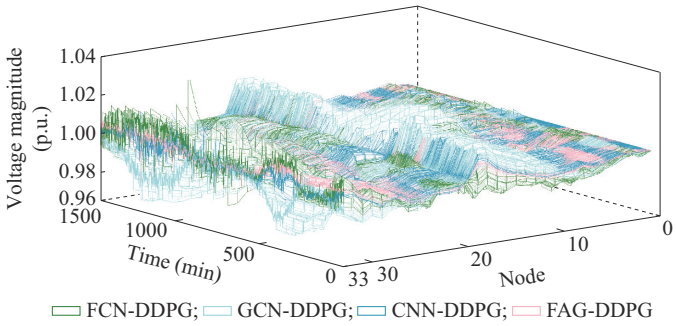

Fig. 12. Voltage profiles of all nodes for different deep learning based models.

TABLE III
VOLTAGE DEVIATIONS FOR DIFFERENT DEEP LEARNING BASED MODELS

Type	Voltage deviation (%)					
	Original	Optimal	FCN-DDPG	GCN-DDPG	CNN-DDPG	FAG-DDPG
Mean	1.66	0.18	0.40	0.51	0.32	0.28
The maximum drop	5.77	1.96	2.31	3.01	3.05	2.25
The maximum rise	1.15	1.80	4.49	2.38	2.95	1.72

C. Application in Real Distribution System

To investigate the performance of the proposed FAG-DDPG model in real distribution system, a modified real 76-node distribution system with seven feeders, four WTs, and two PV systems is employed. The maximum and minimum voltage magnitudes in one episode of testing data with and

without M-SOP voltage control for different deep learning based models are depicted in Fig. 13. The lower maximum voltages and higher minimum voltages demonstrates that the proposed FAG-DDPG model also outperforms other baseline models in the real distribution system. To represent the nodal voltage more intuitively, the voltage magnitudes of node 59 are characterized in Fig. 14. All voltage magnitudes are represented in a three-dimensional (3D) format in Fig. 15, which shows that the voltages are improved by the proposed FAG-DDPG model. Therefore, the proposed FAG-DDPG model has significant potential of voltage control by dispatching the SOP to cope with violent fluctuations of RESs.

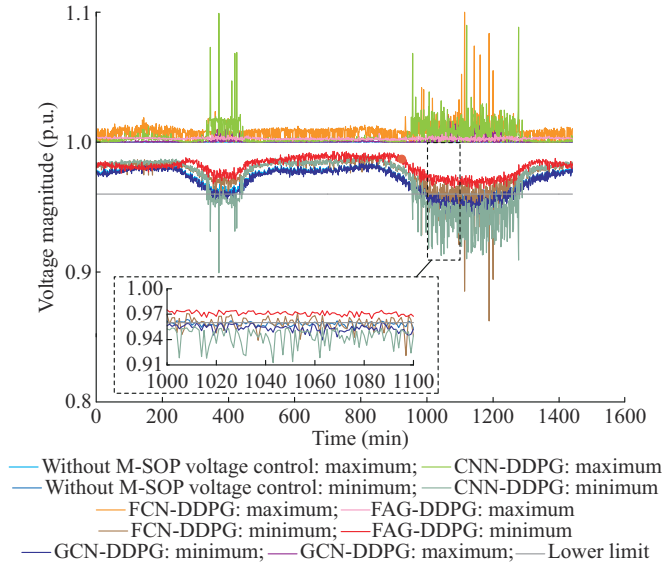

Fig. 13. The maximum and minimum voltage magnitudes in real distribution system with different deep learning based models.

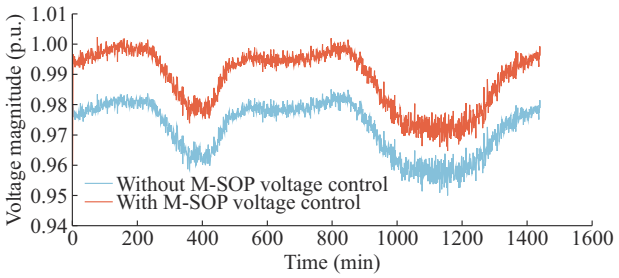

Fig. 14. Voltage magnitudes of node 59 in real distribution system with and without M-SOP voltage control.

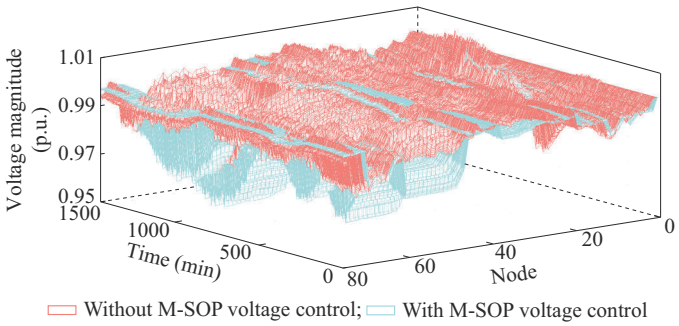

Fig. 15. All voltage magnitudes with and without M-SOP voltage control in real distribution system.

D. Performance with Network Reconfiguration

Since the distribution system topology may change frequently, it is necessary to investigate the generalization performance of the proposed FAG-DDPG model under changing topology. As shown in Fig. 2, there are two tie-lines 8-21 and 25-29. The network reconfiguration may close tie-lines and open other lines. To this end, firstly, the weights of the trained FAG-DDPG model are leveraged to be the initial model parameters. Then, for reconfigured network, the AGCN-based surrogate model and proposed FAG-DDPG model are further trained to adapt to the new topology. Note that the AGCN-based surrogate model is trained to adapt to new topology using previously trained AGCN model and datasets corresponding to the reconfiguration. The maximum and minimum voltage magnitudes under network reconfiguration are shown in Fig. 16.

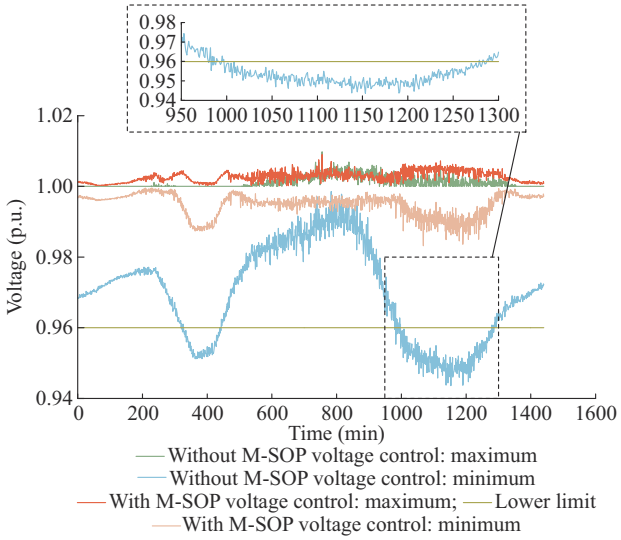

Fig. 16. The maximum and minimum voltage magnitudes under network reconfiguration.

It is shown that the proposed FAG-DDPG model can obtain the optimal SOP control policy with the maximum voltage of less than 1.01 p.u. and the minimum voltage of more than 0.98 p.u. under different topologies. This demonstrates that the proposed FAG-DDPG model has high generalized ability to different topologies with effective performance.

E. Application in Large Distribution Systems

To assess the efficacy of the proposed FAG-DDPG model in large distribution systems, the IEEE 123-node distribution system [30] is utilized as the test system. This system includes four WTs at nodes 13, 25, 62, and 108, and four PVs at nodes 34, 46, 77, and 95, respectively. The M-SOP terminals are set at 32, 68, and 98, respectively. The variations in the maximum and minimum voltage magnitudes, both with and without the M-SOP voltage control, are demonstrated in Fig. 17. The results indicate a marked improvement in voltage stability due to the implementation of the proposed model. Consequently, the propoed FAG-DDPG model exhibits considerable capability in regulating voltage by strategically dispatching M-SOP with rapid fluctuations of RESs.

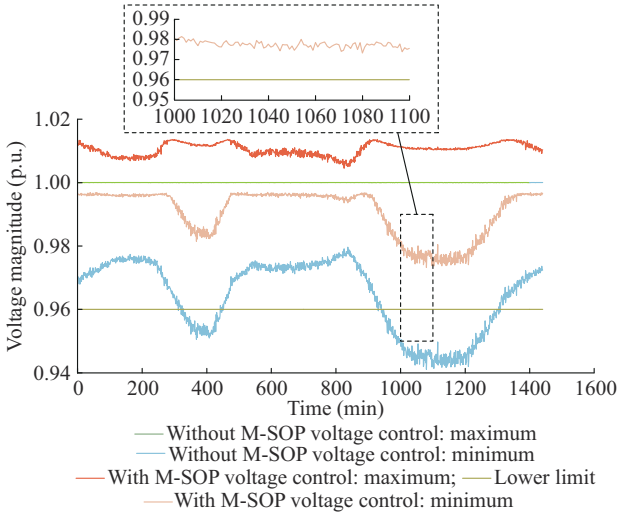

Fig. 17. The maximum and minimum voltage magnitudes in large distribution system.

V. Conclusion

This paper proposes a novel FAG-DDPG model for M-SOP voltage control, which can effectively mitigate the voltage deviation in distribution systems with high penetration of RESs. The proposed FAG-DDPG model with the AGCN-based surrogate model is fully model-free without relying on repeated power flow calculation. The results show that the proposed FAG-DDPG model can adaptively alleviate the voltage deviation caused by the rapid fluctuations of RESs. The superior results on reward and voltage profiles show that the AGCN operation can capture the dynamic graphical nodal features hidden in the environment to facilitate more practical trained voltage regulation policy in enhancing voltage profile. The application of the proposed FAG-DDPG model in a modified real distribution system also illustrates its attractive prospect on M-SOP voltage control against rapid fluctuations of RESs.

References

[1] F. Bai, Y. Cui, R. Yan *et al.*, "Frequency response of pv inverters toward high renewable penetrated distribution networks," *CSEE Journal of Power and Energy Systems*, vol. 8, no. 2, pp. 465-475, Mar. 2022.

[2] Q. Zhao, W. Liao, S. Wang *et al.*, "Robust voltage control considering uncertainties of renewable energies and loads via improved generative adversarial network," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1104-1114, Nov. 2020.

[3] H. Bastami, M. R. Shakarami, and M. Doostizadeh, "A decentralized cooperative framework for multi-area active distribution network in presence of inter-area soft open points," *Applied Energy*, vol. 300, p. 117416, Oct. 2021.

[4] A. Aithal, G. Li, J. Wu *et al.*, "Performance of an electrical distribution network with soft open point during a grid side AC fault," *Applied Energy*, vol. 227, pp. 262-272, Oct. 2018.

[5] J. Li, S. Ge, S. Zhang *et al.*, "A multi-objective stochastic-information gap decision model for soft open points planning considering power fluctuation and growth uncertainty," *Applied Energy*, vol. 317, Jul. 2022.

[6] S. He, H. Gao, H. Tian *et al.*, "A two-stage robust optimal allocation model of distributed generation considering capacity curve and real-time price based demand response," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 1, pp. 114-127, Jan. 2020.

[7] H. Wu, P. Dong, and M. Liu, "Distribution network reconfiguration for loss reduction and voltage stability with random fuzzy uncertainties of renewable energy generation and load," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 5655-5666, Sept. 2020.

[8] I. Sarantakos, M. Peker, N. M. Zografou-Barredo *et al.*, "A robust mixed-integer convex model for optimal scheduling of integrated energy storage-soft open point devices," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 4072-4087, Sept. 2022.

[9] C. Zhang, Y. Xu, Z. Dong *et al.*, "Three-stage robust inverter-based voltage/var control for distribution networks with high-level PV," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 782-793, Jan. 2019.

[10] F. Sun, J. Ma, M. Yu *et al.*, "Optimized two-time scale robust dispatching method for the multi-terminal soft open point in unbalanced active distribution networks," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 587-598, Jan. 2021.

[11] H. Ji, C. Wang, P. Li *et al.*, "Robust operation of soft open points in active distribution networks with high penetration of photovoltaic integration," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 280-289, Jan. 2019.

[12] Z. Liu and L. Wang, "A robust strategy for leveraging soft open points to mitigate load altering attacks," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 1555-1569, Mar. 2022.

[13] Z. Zhang, L. F. Ochoa, and G. Valverde, "A novel voltage sensitivity approach for the decentralized control of DG plants," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1566-1576, Mar. 2018.

[14] X. Chen, Y. Du, E. Lim *et al.*, "Power ramp-rates of utility-scale PV systems under passing clouds: module-level emulation with cloud shadow modeling," *Applied Energy*, vol. 268, p. 114980, Jun. 2020.

[15] Y. Huo, P. Li, H. Ji *et al.*, "Data-driven adaptive operation of soft open points in active distribution networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8230-8242, Dec. 2021.

[16] W. Chen, X. Lou, X. Ding *et al.*, "Unified data-driven stochastic and robust service restoration method using non-parametric estimation in distribution networks with soft open points," *IET Generation, Transmission & Distribution*, vol. 14, no. 17, pp. 3433-3443, Jul. 2020.

[17] R. Leng, Z. Li, and Y. Xu, "Two-stage stochastic programming for coordinated operation of distributed energy resources in unbalanced active distribution networks with diverse correlated uncertainties," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 120-131, Jan. 2023.

[18] T. Ding, Z. Zeng, J. Bai *et al.*, "Optimal electric vehicle charging strategy with markov decision process and reinforcement learning technique," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5811-5823, Sept. 2020.

[19] P. Li, M. Wei, H. Ji *et al.*, "Deep reinforcement learning-based adaptive voltage control of active distribution networks with multi-terminal soft open point," *International Journal of Electrical Power & Energy Systems*, vol. 141, p. 108138, Oct. 2022.

[20] H. Wu, M. Wang, Z. Xu *et al.*, "Graph attention enabled convolutional network for distribution system probabilistic power flow," *IEEE Transactions on Industry Applications*, vol. 58, no. 6, pp. 7068-7078, Nov. 2022.

[21] T. Zhao and J. Wang, "Learning sequential distribution system restoration via graph-reinforcement learning," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1601-1611, Mar. 2021.

[22] Q. Xing, Z. Chen, T. Zhang *et al.*, "Real-time optimal scheduling for active distribution networks: a graph reinforcement learning method," *International Journal of Electrical Power & Energy Systems*, vol. 145, p. 108637, Feb. 2023.

[23] H. Wu, Z. Xu, M. Wang *et al.*, "Two-stage voltage regulation in power distribution system using graph convolutional network-based deep reinforcement learning in real time," *International Journal of Electrical Power & Energy Systems*, vol. 151, p. 109158, Sept. 2023.

[24] Y. Su and J. Teh, "Two-stage optimal dispatching of AC/DC hybrid active distribution systems considering network flexibility," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 52-65, Jan. 2023.

[25] J. Zhao, Z. Tian, H. Ji *et al.*, "Peer-to-Peer electricity trading of interconnected flexible distribution networks based on non-cooperative games," *International Journal of Electrical Power & Energy Systems*, vol. 145, p. 108648, Feb. 2023.

[26] J. Xu, H. Gao, R. Wang *et al.*, "Real-time operation optimization in active distribution networks based on multi-agent deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 3, pp. 886-899, May 2024.

[27] W. Liao, D. Yang, Y. Wang *et al.*, "Fault diagnosis of power transformers using graph convolutional network," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 241-249, Mar. 2021.

[28] T. Zhang, Y. Mu, H. Jia *et al.*, "A successive MISOCP algorithm for

islanded distribution networks with soft open points," *CSEE Journal of Power and Energy Systems*, vol. 9, no. 1, Jan. 2023.
[29] D. Ivic and P. Stefanov, "Control strategy for DC soft open points in large-scale distribution networks with distributed generators," *CSEE Journal of Power and Energy Systems,* vol. 8, no. 3, pp. 732-742, May 2022.
[30] D. Han, L. Wu, X. Ren *et al.*, "Calculation model and allocation strategy of network usage charge for peer-to-peer and community-based energy transaction market," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 144-155, Jan. 2023.

Huayi Wu received the B.Eng. and M.S. degrees in electrical engineering from South China University of Technology, Guangzhou, China, in 2016 and 2019, respectively, and the Ph.D. degree in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, China, in 2022. She is currently a Postdoctoral Fellow in Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University. Her research interests include application of artificial intelligence in power engineering and power system optimal operation with renewables.

Zhao Xu received B.Eng., M.Eng., and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, National University of Singapore, Singapore, and The University of Queensland, Brisbane, Australia, in 1996, 2002, and 2006, respectively. He was an Assistant and then Associate Professor with Department of Electrical Engineering, Technical University of Denmark, Lygnby, Denmark, during 2006-2010. Since 2010, he has been with The Hong Kong Polytechnic University, Hong Kong, China, where he is currently a Professor in the Department of Electrical and Electronic Engineering and Leader of Smart Grid Research Area. He is also a foreign Associate Staff of Centre for Electric Technology, Technical University of Denmark. He has extensive research project experiences involving collaborations with academia, industrial and business sectors. His research interests include smart grid, renewable energy, and applications of artificial intelligence and big data analytics.

Minghao Wang received the B.Eng. (hons.) degree from Huazhong University of Science and Technology, Wuhan, China, in 2012, and the M.Sc. and Ph.D. degrees from The University of Hong Kong, Hong Kong, China, in 2013 and 2017, respectively, all in electrical and electronic engineering. He is currently an Assistant Professor with the State Key Laboratory of Internet of Things for Smart City, Department of Electrical and Computer Engineering, University of Macau, Macau, China. His research interests include power electronics and power systems.

Youwei Jia received the B.Eng and Ph.D degrees from Sichuan University, Chengdu, China, and The Hong Kong Polytechnic University, Hong Kong, China, in 2015 and 2011, respectively. From 2015 to 2018, he was a Postdoctoral Fellow at The Hong Kong Polytechnic University. He is currently an Assistant Professor with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China. His research interests include microgrid, renewable energy modeling and control, electric vehicle, complex network, and artificial intelligence in power engineering.